

Dear Friends,

For over one hundred years, The Rockefeller Foundation has applied new technologies to advance our mission of promoting the well-being of humanity throughout the world. As new technologies mature and proliferate, they spur both optimism and pessimism about their impact on society. When the Foundation funded the Dartmouth Summer Research Project in 1956 – a gathering of renowned scientists, engineers, and mathematicians who coined the term “artificial intelligence” (AI) – the concepts driving AI were so new that their implications were understood only by the innovators who were developing them. Today, with further development and integration into society, AI’s impact on our lives is much more consequential.

The Rockefeller Foundation continues to harness the frontiers of data, science, and technology for the good of all humanity, particularly in service of catalyzing a more inclusive, green recovery from the Covid-19 pandemic. Whether an algorithm can predict the emergence of a potential future viral threat or more precisely inform demand for energy access, we believe AI applications have the potential to help us build a more resilient post-Covid world. But we are also conscious of the associated risks and the need to mitigate downsides.

In October 2019, The Rockefeller Foundation convened an exceptional group of technologists, philosophers, economists, lawyers, artists and philanthropists at its inspiring Bellagio Center to discuss AI’s role in creating a better future for humanity. The report [AI+1](#) captured some of the ideas from that convening, including a strong consensus on the need for AI governance.

The world changed dramatically soon after that convening and AI has played an important role in advancing responses to Covid-19, as well as a growing role in the longer term recovery. How we govern AI’s development and application will strongly influence whether the recovery from today’s crises helps resolve the social, economic, and environmental injustices revealed by Covid-19 – or whether it will exasperate them further. To realize AI’s potential for a positive influence, we must accelerate progress on AI governance. We must move beyond high-level principles and ethical guidelines towards new ideas, solutions, and practices.

There is no one-time fix. A new AI governance “field” is needed to address the problems of today and those we cannot anticipate tomorrow. Field-building involves shaping new ideas, creating communities, and developing new practices and institutions. We’re delighted that you’ve joined us in this endeavor.

Our first step together, in the convening on December 4th, will be to identify the most important research questions and shape new ideas and frameworks for AI governance. Six pre-read papers will establish a common and shared baseline level of knowledge, covering essential concepts and historical

context, as well as practical applications and analogies that will stimulate our thinking on new strategies for how to design and deploy effective governance frameworks. One set of readings provides a level set of basic concepts, historical antecedents for AI governance, and trustworthiness in the context of AI technologies to inform our first session together. The second set covers global regulatory institutions, current initiatives in AI governance, and a deeper dive into the concept of regulatory markets to inform the second session. Please read these materials in advance of the convening, as we have designed our time together to minimize presentation and maximize discussion.

The first session of the convening will be facilitated by Stanford University's Center for Advanced Study in the Behavioral Sciences. We will cluster in small breakout groups where participants will explore and shape the research questions of governance addressing underlying issues of human-machine compatibility that are relevant to AI's long term development and near-term pressing challenges. We request you complete this [pre-convening survey](#), which will inform how we facilitate and structure these initial sets of breakout discussion groups.

After the first session, participants will have time to reflect, gather thoughts, or virtually connect 1:1 with other participants during an unstructured break period via coffee chats.

The second session will be facilitated by the University of Toronto's Schwartz Reisman Institute. They will introduce a solution design approach intended to translate ideas surfaced during the morning discussion into practical applications. Coming out of this session, we will identify an initial portfolio of projects that would benefit from integrating specific governance frameworks in order to inform future field-level infrastructure needs.

Our ambition is to collectively identify the most critical research questions that will advance the field of AI governance, and ultimately pressure test bounded projects to shape solutions with real-world impact. You bring diverse perspectives and unique knowledge, which is why your participation will be so valuable, and for which we are very grateful.

We look forward to you joining us.

Cheers,



Zia Khan
Senior Vice President, Innovation